DOCUMENT RESUME

ED 091 903 EC 062 127

AUTHOR	Goodstein, H. A.
TITLE	The Role of Research and Evaluation in EMH Curriculum Development: Project MATH.
INSTITUTION	Connecticut Univ., Storrs.
SPONS AGENCY	Bureau of Education for the Handicapped (DHEW/OE), Washington, D.C.
PUB DATE	Apr 74
GRANT	OEG-0-70-2250(607)
NOTE	11p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974)
EDRS PRICE	MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS	Curriculum Development; *Curriculum Evaluation; *Educable Mentally Handicapped; Evaluation Criteria; *Exceptional Child Research; Field Studies; *Mathematics; Mentally Handicapped; *Research Design; Research Methodology

ABSTRACT

This paper reviews evaluation and research activities of Project MATH (Mathematics Activities for Teaching the Handicapped) during 1972-1973, and discusses evaluation of curriculum materials for educable mentally handicapped (EMH) populations. It briefly describes field tests in six cities involving our 100 teachers (primary through junior high levels). The field tests, concerning number and operation strands, involved program evaluation, collection of biodemographic information, teacher tracking of daily instruction, computer processing, and a questionnaire for teachers. The paper also describes concurrent curriculum review, another review by mathematicians, and implementation of research studies. The following issues are considered: (1) Researchers tend to make evaluation designs that overestimate the amount of data necessary for revision. (2) Researchers tend to overestimate usefulness of empirical data for curriculum evaluation. (3) The "representative" field test has a hallowed position it may not deserve. (4) The nature of the educable mentally retarded population restricts usefulness of evaluation designs that rely on pupil change data. (5) Demonstration of effectiveness decreases with increased magnitude of the curriculum being developed. It is suggested that many issues thought to require large-scale field tests could be determined with a few carefully controlled research studies, and that mechanisms and criteria be developed to select the best sequence for studies. It is noted that Project MATH people have decided to stay with small sample, short duration studies. (MC)

Session - 19.13

ED 091903

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE ROLE OF RESEARCH AND EVALUATION IN EMH CURRICULUM DEVELOPMENT: Project MATH

A Paper contributed as part of a Symposium to be presented at the Annual Convention of the American Educational Research Association, Cosponsored by the Special Interest Group Special Education Research, April, 1974.

H. A. Goodstein
Associate Professor of Education

University of Connecticut
Storrs, Connecticut 06268

*This paper was prepared pursuant with a grant from the Bureau of Education for the Handicapped, USOE, "A Program Project Research and Demonstration Effort in Arithmetic Among the Mentally Handicapped" OEG-0-70-2250(607) #162008, under the direction of J. F. Cawley, Project Director, H. A. Goodstein Assistant Director, University of Connecticut, Storrs, Connecticut 06268.

EC060212

The Role of Research and Evaluation in EMH Curriculum Development:

Project MATH

Project MATH (Mathematics Activities for Teaching the Handicapped) has as its mission the development and validation of a mathematics curriculum for handicapped children. The curriculum model adopted for Project MATH provides for the slower rate of cognitive development of many handicapped children, provides diagnostic alternatives to instruction, and provides for the affective and behavioral growth of the child. The curriculum is considered multiple-option in that it seeks to provide teachers an array of instructional and content options necessary for teaching youngsters who have failed in the traditional curriculum.

Evaluation and research are two of the five components of Project MATH (the remaining components being development, training, and dissemination). Evaluative activities of Project MATH (to this juncture) have been restricted to internal and external interim formative product development (Sanders and Cunningham, 1973). Research activities have been related to the examination of issues in assessment of needs and to the performance of handicapped children on tasks which are analogies of potential curriculum activities.

Briefly, the following sections of this paper will relate the management of evaluation and research activities of Project MATH during the 1972-73 academic year. Following this section, a discussion of issues relating to these areas will be attempted.

Two strands of directed teaching activities were completed and ready for field testing by September, 1972. These strands were numbers and operations (primary, intermediate, and junior levels) and sets and operations (primary level only). Each strand included more activities than could be taught during any seven months. For each arithmetic topic, a range of activities were included that differed along the dimension of teacher-pupil instructional interaction.

The purposes of the evaluation effort were: (1) to determine the percentage of teachers who would employ the lesson guides on a regular basis; (2) to determine the sequences of activities chosen by the teachers; (3) to determine the number of activities taught during the instructional day; (4) to determine the teacher ratings of pupil performance on each lesson; and (5) to determine the teacher ratings of the quality of each lesson guide.

Each classroom to which Project MATH materials were distributed furnished a list of the children with the following biodemographic information: Age, sex, race, IQ, administrative classification of handicap, and parental occupation (this variable was not used because of lack of data). Additionally, each classroom was described as either self-contained or resource room.

Teachers tracked their instruction on a daily basis. Each activity taught to a child was recorded. All lessons were evaluated along two dimensions: teacher judgment of pupil performance on a three point continuum (failure, learning, and mastery) and teacher judgment of the quality of the lesson as written (good, adequate, or poor).

Data sheets were collected on a regular basis, keypunched, and entered into a master computer data file. Additionally, an evaluative questionnaire composed of Likert-type items regarding the curriculum was sent to all teachers near the end of the project year.

In addition to the field testing of the curriculum, internal review of the curriculum proceeded even after the distribution of the materials. This review centered upon the consistency of the lessons to the development model, adequacy of the directions, and clarity of the mathematics content. As a supplement to this review process, an advisory board consisting of professionals in mathematics education and special education were asked to respond

in writing to the adequacy and clarity of the instructional materials.

The research program was developed independently from the evaluation process. A series of research studies were designed by project personnel as a means of providing input into future decisions regarding development of materials or instructional tactics. Studies were proposed to the project management and reviewed for significance of potential data for decision-making, availability of necessary resources, and design clarity. An independent data acquisition team was recruited to collect data for the studies.

The proposer of the study assumed responsibility for procurring the necessary testing materials, training the data acquisition team in the experimental procedures, monitering the experimental procedures, designing the data analysis systems, and interpretation and writing the project report. Subjects were acquired from school systems local to the project, many of which were not involved in field testing of materials.

The use of a data acquisition team independent from the regular project staff allowed the project to conduct approximately twelve studies without significantly diverting human resources from the curriculum development efforts. For each week a research study was operative, only one staff member was engaged in the monitering of that study.

II

Having completed a full year of managing both a moderate scale field test (over 100 teachers in six cities) of instructional materials and a continuous research program, it is useful to reflect upon issues relating to research and evaluation in curriculum development for handicapped populations. The conclusions or implications drawn represent solely the opinion of this writer and should not be judged as representing the thinking of other project staff members or the funding agency.

(1) Evaluation designs developed by researchers tend to over-estimate the amount of data necessary to the formative process of revision. The necessary corollary to this first generalization is that necessary formative information must then be abstracted from a wealth of data, diverting time and other resources away from the use of the necessary data in the revision process. A basis decision must be made as to how much resources should be earmarked for collection of ancillary data, not directly applicable to the revision process, whose ultimate value will rest as a data base to test research hypotheses regarding the use of the curriculum.

The temptation is all too real to collect data from an available population on an array of instructor and learner characteristics. Each pieces of data can be justified in terms of legitimate research hypotheses, explicated or potential. However, the aggregate effect of this data collection process may result in obscuring of the focus of the evaluation effort. One should clearly deliniate that aspect of the data pool which will yield direct formative payoff and develop procedures that maximize the acquisition and analysis of that information.

(2) We tend to over-estimate the usefulness of empirical data in the process of formative evaluation of curriculum. Research designs rely heavily upon empirical sources of data. Evaluation designs, often developed by researchers, may tend to underestimate the collection of the non-empirical: teacher comment and revision of lessons, expert logical or rational analysis of materials, child reaction to the curriculum.

It could be that we spend most of our resources on collection of empirical data, but rely most heavily on the "soft" data for ultimate revision decisions. Our uncomfortableness with such data sources may reflect our training more than an objective analysis of the usefulness of such data sources.

(3) We have elevated the "representative" field test to a hallowed position it may not deserve. The literature in curriculum evaluation presumes that materials should be field tested as a component of formative evaluation. Further, the objective of the field test is to demonstrate or fail to demonstrate pupil change in response to components of the program. Failure to demonstrate change becomes the stimulus to the revision process.

Field testing is an expensive process. Materials must be printed and disseminated, teachers must be given some minimum level of in-service training, site visits must be conducted, and empirical determinations of pupil progress made. The revision cycle for those materials must be delayed often for the entire duration of the field test and for the necessary time for data reduction, analysis, and interpretation.

The fact is that this entire set of procedures comprises an act of faith. To my knowledge there is no research that indicates that field testing a product to determine formative revision on the basis of pupil change data results in a more effective product in a summative sense. Here, one is not refering to pilot tryouts of materials for the purpose of collecting criticism and revisions of the materials, but rather the more large scale field tests we are all familiar with.

(4) The nature of the EMR population restricts the usefulness of evaluation designs that rely upon pupil change data. Generally, EMR children enter the special education system after a number of years of unsuccessful adaptation to instruction in the regular grades. Epidemological surveys indicate little special education placement prior to the third grade. Thus, from the outset, special education is forced into a remedial as well as a developmental mode in response to the education of the EMR. Rather than expecting a learner to be intact in regard to necessary prerequisite skills and understandings, it

is more likely that many children will have specific learning disabilities.

For the average child, one might be able to project an expectation of one year's growth in grade level achievement for each year the child is exposed to a developmental curriculum. If the EMR population was characterized as intact and approximately placed in a developmental curriculum, one might project growth expectancies in relation to average IQ. That is, if the population has a mean IQ of 75, a growth rate 3/4 of a grade equivalent could be projected. However, in a population characterized by failure sets, inappropriate development of prerequisite skills, and large experience gaps, what standard should become the criterion for judgement of adequacy?

Does requirement of a criterion-referenced framework for a norm-referenced framework resolve this difficulty? Rather than looking toward a grade equivalent standard, we shift our emphasis to a series of program objectives. The mastery of 'x' number of objectives becomes the program goal. However, we are still forced into making some a-priori decision regarding the number of objectives mastered that we are willing to accept as a criterion of adequate progress. If the system lacks this a-priori standard, the process of evaluation becomes totally descriptive without a judgemental component.

Accordingly, we in fact impose normative expectations upon our criterion-referenced measurements. This is where the system causes difficulty for judging EMR populations responses to a curriculum. Divergent achievement patterns often may be the rule rather than the exception. What judgement does one make of the efficacy of a curriculum where only a minimum number of objectives are mastered?

The process of judging the adequacy of an instructional sequence by achievement also deserves closer scrutiny because of inherent population characteristics. In this model, failure is most often presumed to lead to revision

of the instructional materials. However, unless the instructional program has identified all specific prerequisite learning sets and task requirements (a dubious assignment in view of our present limited knowledge base), it may be that failure is a function of a lack of readiness on the part of the child.

Instructionally for the single child, we accommodate a wide range of individual differences by specifically allowing for alternative instructional strategies or alternative objectives. However, evaluation designs have traditionally been limited to group data decisions. Achievement of objectives is evaluated in terms of group average achievement standards. An objective which shows a low group average achievement may not need revision of the instructional program. The instructional program may have been successful for those students who were ready for program objective and will be successful for other students an another juncture in the instructional program.

(5) The larger the magnitude of the curriculum being developed, the less chance of demonstrating effectiveness through the formative process. The truth value of this statement will be modified by the length of time allowed for this evaluation. The shorter the duration allowed for the evaluative process, the less likelihood of demonstrating effectiveness (that is, by the usual criteria of empirical data relating to achievement). Additionally, the older the age of children for whom the curriculum is introduced, if it is for a developmentally organized subject matter, the less likely effectiveness can be demonstrated.

A developmentally organized curriculum in mathematics or reading, which is so heavily dependent upon some sequential order of skill mastery and spans a wide range of chronological age development, relies upon cross-sectional evaluation

designs. That is, unless one was willing to wait six or more years for a longitudinally organized formative evaluation, it must be introduced with children at different age levels, and with children who were not previously exposed to the curriculum. This maximizes the impact of pre-requisite skill lags or omissions and minimizes the probability of demonstrating effectiveness.

The feelings of this writer is that we must critically examine previous assumptions regarding what constitutes adequacy of the formative evaluation process. This writer would argue for more reliance upon review of materials and instructional programs by various levels of "expert" opinion — subject matter specialists, educational psychologists and/or special educators, teachers, and administrators. Better review procedures and research on the process of review should be developed.

Field testing of the materials could be viewed as a component of this review process. Data on teacher ease of implementation, attitudes toward the program, changes in the intended instructional activities, and specific pupil performance difficulties encountered should be collected. Once we separate the field test from an inherent research model to a more purely evaluative model, the requirements for large, "representative" field test populations could be reduced. One might then opt for more intensive study of smaller number of teachers, who were motivated to fully participating in vigorous formative review.

III

From its inception, Project MATH has invested heavily in a research component separate in its organization from the evaluation process. Most of the research studies have used small, carefully selected samples of children to test carefully structured curriculum-related hypotheses. A large majority

of these studies have focused upon verbal problem solving processes. The impact of many of these studies is directly reflected in the curriculum, specifically the verbal problem solving component.

Overall, the pay-off from the research investment can be judged as considerable. However, this pay-off was less than maximal due to a degree of independence given to the proposers of the research endeavors, independence from the priority of research needs mandated by the development efforts. It probably is a more maximizing arrangement to have research studies organized from questions raised from the development process, rather than relying upon the probability that independently generated research studies will have an impact upon the development process. Perhaps, many issues we have attempted to resolve through the large-scale field test could be determined through a limited number of carefully controlled research efforts.

A potentially useful area of future inquiry is what management mechanisms and decision criteria must be developed in order to select the most heuristic sequence of research studies to be completed. Our project has elected to stay with small sample, short duration studies. Whether such a research program organization is the most effective is open for discussion.

This paper has attempted to develop several issues that should and must be discussed by those interested in the process of curriculum development and evaluation for handicapped populations. Some of these issues may even extend to curriculum development for "average" learners, to the extent that individual differences may prove the universal phenomenon in education.

References

Sanders, J. R. and Cunningham, D. J. A structure for formative evaluation in product development. Review of Educational Research, 1973, 43, 217-236.